

Identifying social network effects

Article (Accepted Version)

Dolton, Peter (2017) Identifying social network effects. *Economic Record*, 93 (s1). pp. 1-15. ISSN 0013-0249

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/67383/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Identifying Social Network Effects*

Peter Dolton

(University of Sussex.)

Abstract

This paper reviews the current state of empirical econometric identification in the economics of networks. Possible identification strategies which exploit the properties and characteristics of networks are described. The main arguments are illustrated with examples from the US AddHealth data and an account of the Bloomsbury Group network. Prescriptive suggestions, based on the way in which networks actually form and operate, are made for a more considered approach to empirical econometric work which involve networks.

Address for Correspondence:

Prof Peter Dolton
Department of Economics
University of Sussex
Brighton, BN1 9SL
Sussex, UK

Email: Peter.Dolton@sussex.ac.uk

- This paper is based on the plenary lecture given at the ACE conference in Adelaide in July 2016. I wish to acknowledge the research assistance of Lucia Barbone in the data preparation of this paper.

Acknowledgements: Part of this paper uses AddHealth data. This is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data _les is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

1. Introduction: Why are Networks Important?

Nearly all individuals have family, friends, colleagues and acquaintances. As consumers we all undertake transactions with shops, online stores and other providers. Likewise firms, countries, clubs, pressure groups, political parties, trade unions, are all inextricably related and connected to one another, and other agents, to a greater or lesser degree. Networks pervade all social, political and economic interactions. The functioning of markets, organisations, crowds and relationships are all governed by network structures of different types. These network structures have spill-over effects and externalities, both on their own members, and those outside the network. As economists, the modeling and calibration of network effects is therefore of central importance to understanding individual behavior and the functioning of markets. At the level of the empirical econometric model this means that each unit of observation is not independent of other observations in the dataset. As a consequence each row of the data could be related to other rows of the data and possibly the variance-covariance structure of the unobserved heterogeneity could be non-diagonal. Therefore, the categorization and econometric identification of network effects could be complex. Here we provide a descriptive overview of these effects and the problems applied econometricians face in trying to estimate them. We focus on a descriptive introduction to empirical network econometrics and we use concrete examples to show: firstly, how real networks may be more complex than econometricians have so far considered; secondly, we introduce an example application in which the interpretation of what network effects actually are can be questioned, and thirdly we provide a simple example of a way in which networks form and can be described may be very different from what we currently assume.

Social networks influence and change behaviour. The extent to which our decisions and outcomes are dependent on: which network we belong to; how many members the network has;

and what our position in the network is, can be the source of great variability. In short, networks create and modify the effect of spillovers and externalities. Any person in a network may have their key outcome determined partly by the outcomes of others in their network. For example, it might be that I am more likely to be obese if my friends are obese. (See Christakis and Fowler 2007). These are what economists like to call *endogenous network effects*. Likewise, it could be the case that my outcome could be affected by my friend's or peers characteristics. For example, I could be more likely to be obese if my friends eating patterns include fast food (as I might be likely to accompany them). These are called *exogenous network effects* by economists.

Economists have considered that many externalities could be partly a function of network contacts and interaction. In labour markets (Myers and Schultz, 1951, Rees and Schultz 1970, Ioannides et al 2004 and Goyal 2007) there is a focus on the role that contacts play in getting jobs. In criminal behavior (Reiss 1988, Glaeser et al 1996) economists are worried about endogenous network externalities which show how criminal behaviour is learnt from those we associate with. Likewise in family decisions (Fryer 2007, Rainie and Wellman 2012), and health behavior (Christakis and Fowler, 2007, 2008, 2011) we are worried that problems occur as spill-over effects from those close to us. Development economics (Fafchamps and Lund, 2003, De Weerd 2004), financial decision making (Elliott et al 2012), and agricultural and technological adoption involve diffusion processes (Griliches 1957, Coleman et al 1966, Vega-Redondo 2007) which are determined by network phenomena. The role of information transmission between agents in auctions and the consequences of the world wide web are complex, as are, the implications of networks for organizations more generally (Uzzi 1966, Weisbuch et al 2000, Barabasi 2002, Pentland 2014).¹

¹ For those seeking an introduction to the subject the starting point would be (in order of difficulty): Knoke and Yong (2008), Borgatti et al (2013), Kadushin (2012), Newman (2010) and Wasserman and Faust (1994).

The impact of exogenous changes or policy interventions may have effects that are different because of the existence of networks. Likewise, different members of the network are affected differently by such changes, as a result of their different position in the network. Any individual acting alone may not be greatly affected or change their behavior as a result of some policy or exogenous event. But, because we are influenced by our friends, we may be stirred into action. If we want to measure the aggregate impact of any change (including externalities) in networks then we are interested in what is the ‘*social multiplier*’ effect of some policy change or exogenous event.

Information passes through networks at different rates depending on how: connected the network is; what the processes of diffusion looks like; and how the contagion may work. Of course there could be a huge degree of variation in how many contacts each individual has. This means the role that any individual may play in this information diffusion will vary considerably. We could naturally be interested in which agents in the network are correspondingly so relatively important. Identifying who these ‘*key players*’ are depends on being able to measure reliably how many contacts a person has and whether they are reciprocated. We would also wish to calibrate the extent to which each node is at the centre of the network. Such a measure of ‘centrality’ is core to measuring the influence of each person in the network: which will be a measure of how many contacts or friends, and friends of friends, and friends of friends of friends, and so on, that a person has. Separately, one should be interested in the extent to which, if one is central to a network, whether this impacts on their own outcomes. We will call this the ‘*key player effect*’ and return to the measurement of this effect of this definition in the next section.

It is clear that economic theorists have contributed hugely to our theoretical understanding of strategic interactions between network members by treating them ‘as if’ they were players in a

game (see Vega-Redondo, 2007, Goyal, 2007, Jackson 2008, Easley and Klienberg 2010 and Blume et al, 2011 for overviews of this literature). This paper will argue that applied econometricians need to distinguish carefully between different types of network effects and devote more attention to their empirical estimation. To this end we first outline what these econometric identification problems are. We next provide a concrete simple empirical example of how a network is formed and may be described by looking at the Bloomsbury Group. In section 4 we take a look at data from the AddHealth data to show how simple theoretical models of network structure can be wrong in their empirical predictions. In concluding we outline some of the biggest challenges for empirical econometric work involving networks.

2. The Network Model: Key Empirical Econometric Questions.

The standard ‘Linear in Means’ peer effects model is the appropriate place to begin a description of the econometric issues of modeling network effects and its attendant identification problems.

This section is a simplified recap of the main results in Manki (1993), Moffitt (2001), Bramoulle et al (2009), Calvo-Armengol et al (2009), Blume et al (2011), Angrist (2014), Advani and Malde (2014), Angrist (2014) and Boucher et al (2016) for the empirical researcher.² Here we focus on the most substantive results in terms of their implications for delineating what network effects there are and how they may be identified.

One can write the basic form of the ‘peer effects’ or Linear in Means model (Manski, 1993) as:

$$y = \alpha + \beta Gy + \gamma x + \delta Gx + \epsilon \quad (1)$$

Where y is some outcome of interest, written in vector form, for K individuals. The vector x is

² The interested reader is referred to these references for proofs of the main propositions we describe below and the details of related theoretical propositions. An alternative exposition which provides the code to analyze a dataset is in O’Malley and Marsden (2008). Recent comprehensive theoretical summaries of econometric identification in networks can be found in Blume et al (2011) and Chandrasekhar (2016) .

some exogenously determined characteristic of these individuals³ and G is the adjacency matrix characterized by zeros and ones to indicate non-connection and connection between individuals respectively. We also assume that $E(\epsilon|x, G) = 0$. This assumption implies that the x and G are exogenous to the determination of unobserved heterogeneity. Such an assumption is a major limitation of the model. Note also that for estimation, some assumption needs to be made about $E(\epsilon\epsilon')$. Given the structure of the model it is unlikely that it is realistic to assume that $E(\epsilon\epsilon') = \sigma^2 I$. This in turn implies that $E(yy')$ will have a non-standard form and this could pose additional problems. Alternatively certain kinds of variance-covariance restrictions could provide opportunities for identification (See Graham, 2008). This provides a logical link to the spatial econometrics literature in which some specific form is assumed for the Variance-Covariance structure – usually geographical contiguity or some other instrument. (See Anselin, 1988 for a full treatment or Dolton et al, 2016 for an example.)

This model suggests that each person's outcome could be a function of the average of others outcomes to whom they are connected and possibly to the average of other's characteristics. The former effect, captured by β , is called the *endogenous effect*. The latter effect, captured by δ is the *exogenous effect*. This model has been used by lots of authors to attempt to capture – so called – ‘peer effects’ by simply including the mean values of peers y 's and x 's as regressors in their outcome equation. Regrettably it is usually the case that these network effects are either not identified or it is hard to interpret exactly what is being measured by their estimation.

A little algebraic manipulation of (1) gives us:

$$y = \alpha(I - \beta G)^{-1} \iota + (I - \beta G)^{-1}(\gamma x + \delta Gx) + (I - \beta G)^{-1} \epsilon \quad (2)$$

But it is useful to note that matrix in brackets (to be inverted) can be approximated by:

³ Note that x could be a set of characteristics with no loss of generality.

$$(I - \beta G)^{-1} = \sum_{k=0}^{\infty} \beta^k G^k \quad (3)$$

Where it should be noted that the exact strict upper bound for the scalar β is given by the largest eigenvalue of the matrix G . This is an important and convenient result as it can be used to define the Katz-Bonacich measure of network centrality. (see Bonacich 1987)⁴

Which means that a reduced form of (2) can be rewritten as:

$$y = \frac{\alpha}{(I-\beta)_i} + \gamma x + (\gamma\beta + \delta) \sum_{k=0}^{\infty} \beta^k G^k + \sum_{k=0}^{\infty} \beta^k G^k \epsilon \quad (4)$$

The expected mean friends group outcomes conditional on x is then:

$$E(Gy|x) = \frac{\alpha}{(I-\beta)_i} + \gamma Gx + (\gamma\beta + \delta) \sum_{k=0}^{\infty} \beta^k G^{k+2} x \quad (5)$$

In the context of this model (or logical sub-cases of this model it is possible to prove the results below:

Proposition 1. (Manski 1993) The Reflection Problem. If γ and $\delta=0$ then $y = E(y|x) + \epsilon$. Which means that we can say nothing about endogenous social interactions.

We cannot distinguish the effect of one individual i 's outcomes on others, from their other people's impact on i or anyone else. Intuitively, if everybody in the sample affects the mean, then how can we simultaneously determine the effect of the mean on each person?

Proposition 2. (Manski 1993) The Identification Problem. If $\beta \neq 1$ and $\delta=0$ then it is not possible to simultaneously determine α , β and γ in a structural form representation of equation (1).

This is essentially why the peer effect literature is so problematic. More details of the empirical problems of identification in peer effects models are provided by Angrist (2014).

⁴ It is convenient that this measure approximates the paths of length k for all nodes in the adjacency matrix – which is the essence of the centrality measure capturing the notion of friends of friends ($k=2$), friends of friends of friends ($k=3$) and so on.

Proposition 3 (Bramouille et al 2009) Suppose $(\gamma\beta + \delta) \neq 0$. If matrices I , G and G^2 are linearly independent, network effects are identified.

Proposition 4 (Bramouille et al 2009) Suppose that individuals interact in groups. If all groups have the same size, social effects are not identified. If (at least) two groups have different sizes and $(\gamma\beta + \delta) \neq 0$, social effects are identified.

The corollary of Proposition 4 is that use can be made of small heterogeneous groups with local differences or non-overlapping groups can be used for identification (see DeGiorgi et al 2010 for example). Likewise, block structures (for example a dataset like of AddHealth which consists of pupils from different, non-connected, schools) can also be used for identification.

Proposition 5. (Calvo-Armengol et al 2009) Assuming agents in a network choose effort levels of inputs in a simple quadratic way then the Nash Equilibrium in the network involves individual outcomes which can be uniquely defined in terms of Katz-Bonacich centrality.

This result is simple but powerful. It suggests that under simple, fairly plausible, regularity conditions the outcome equation across a sample of network members can identify a term which is a function of the individual's centrality to the network. So this regressor – in effect – captures the importance of that person's position in the network - to their own outcome – what we have called the 'key-player effect'. Notice that this is not the same thing as the other effects of networks that we would also like to identify. Specifically, it tells us nothing about either the endogenous or exogenous effects of other network members on the individual – nor does it enable us to identify any social multiplier of any exogenous change. Rather, it tells us if being at the heart of a network – in a well defined sense – has any explanatory power in determining outcomes. In many contexts this could be a very informative question.

So far our description of the results in the literature has ignored the most difficult problem.

Namely, how do we proceed when the G adjacency matrix is endogenous – so that the process by which people form links to other members in a network is determined by unobserved heterogeneous factors – like personality, charisma, energy, drive, enthusiasm, sense of humour, and other character traits – which themselves may also be important in the determination of any outcome of interest. In this situation it will be potentially difficult to determine effects which are due to the true endogenous (or exogenous) effect of networks rather than to the process of the formation of a network – how can I be sure that I have estimated the endogenous effect of having obese friends, on my obesity, when the impact could really be down to the fact that I hang out with people like myself in terms of personality and outlook on life and they just happen to be obese.

Various authors have proposed different ways to tackle the endogeneity of G . Bramoulle et al (2009) suggest that: local differences in parts of the network; block differences in the network and instrumental variables could all provide solutions in different contexts. Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016) propose ways of structurally estimating the component elements of the adjacency matrix and then, having done this, estimate the underlying model of outcome determination. The validity of this strategy usually relies on having exclusion restrictions for the friendship formation process that play no role in the outcome determination. Examples could be various types of homophily or propinquity that can be argued are exogenous. Goldsmith-Pinkham and Imbens (2013) and Boucher et al (2016) go one stage further by providing suggestions for how the applied researcher may test for endogeneity of G .

It should be observed that this standard exposition of the network estimation model is that it does not capture: any dimensions to links (it is ‘as if’ each link was a unidimensional unit of connectedness), or the intensity of links (the standard default assumption is that all links have

the same weight), or what happens when links are not reciprocal. Nor does it tell us how a network is formed or how it develops over time. It is also silent on how the presence of homophily - the degree of similarity between individuals - or propinquity – how spatially close individuals may be located - in the network may condition the endogeneity of the adjacency matrix. In addition, there is no clear definition of what exactly the social multiplier is in this model or how we should measure exactly what constitutes a key player in a network. One could suggest that a measure of centrality could be used but then we will show below that we don't really have a satisfactory – empirically verifiable theory – of how network connections are made.

What are the aggregate spillover effects that occur because of networks? Can we identify them? What is the Effect of a Network (or network position) on an outcome for the Individual? Who is/are the Key Player/s – Can we identify these people (agents, firms, countries etc) and how much does it matter to them? These are all important empirical questions which we would variously like to seek answers to.

In the estimation of an outcome equation for individuals, what might be the biases of ignoring networks connections – assuming we had mistakenly assumed cross section independence? How tight or clustered are networks and does it matter for processes and outcomes? What role do networks play in diffusion and contagion? More generally how do networks form and change over time – what are their dynamic properties? What do we learn from network connections? What is the distribution of the number of contacts or nominations? Does the theory match the reality of what distribution we can observe empirically?

One area that economists have devoted huge attention to is the role to strategic factors play in above processes. The theory chapters in Jackson (2008) and Easley and Kleinberg (2010) are devoted to overviewing the technical papers which answer questions relating to the effect of

strategic interaction between individuals in a network, as if they were rational players in a game. These papers are predominantly theoretical and devote modest attention to the empirical assessment of their theoretical results. Often when economic theory does confront the data it is to ask fairly limited questions with very simplified concepts and ‘toy’ data sets, which ignore much of the complexity we have described above. Our purpose in the next section is to describe a real life network in some detail to see if we can address any of these puzzles.

3. An Example: The Bloomsbury Group

We do not have very many good accounts of how networks form and how their diverse structures may evolve over time⁵. The age of the great diarists and correspondence is largely over. The age of email, written in the virtual ether and lost as computer backups are erased, does not leave a trail for us to follow. Those letters, memoirs and diaries from the last century are in our great libraries (although slowly being consigned to store houses and archives – never to be consulted) but the emails are essentially private and rarely seen by other than those for whom they were meant (or copied to). Blogs are what have replaced these interactions – but for the most part blogs are not reactive – they are simply one person’s take on the world. One well documented network formation and interaction which provides an intriguing insight is the collective machinations of the Bloomsbury Group which has been well documented. (See Edel 1979, Bell 1968, 1995, Spalding 2005). This group which began

Figure 1. The Bloomsbury Group Network circa 1905.

⁵ One notable, well cited exception which has been used to reveal many insights into network structure and implications is the work of Padgett and Ansell (1993) on the connections between the Renaissance Florentine families.

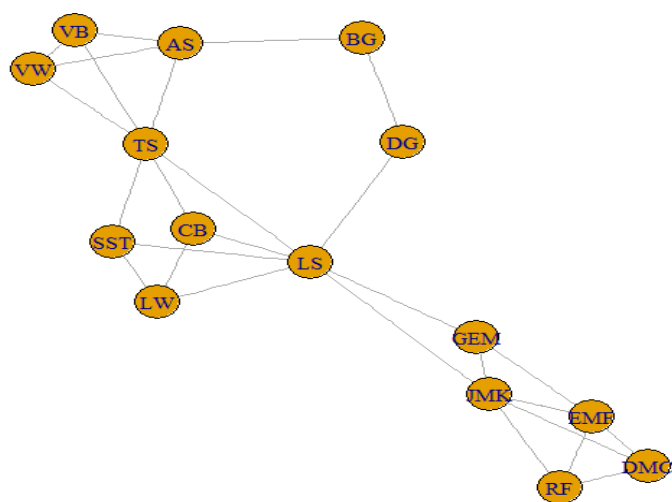
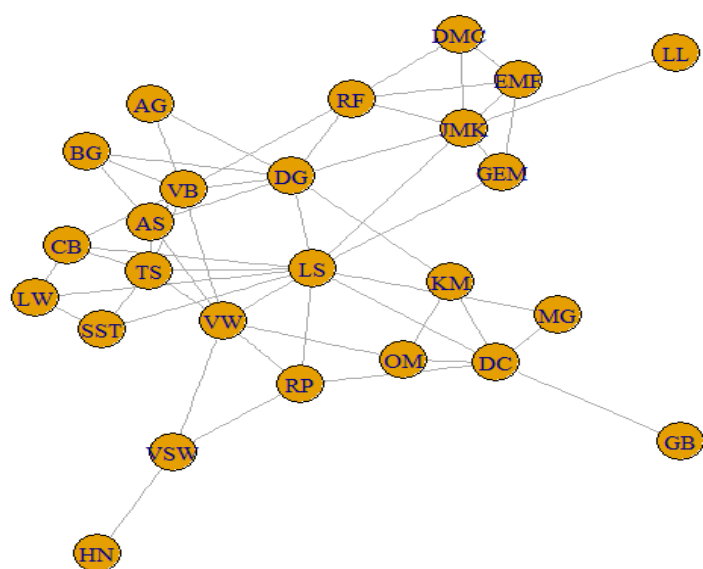


Figure 1. The Bloomsbury Group Network circa 1925.



in 1904 and lasted until 1938 provides a fascinating insight to networks, their development, their complexity, and their possibilities – all of which are especially relevant to economists as John Maynard Keynes was at their heart. There is also understandable interest in the way in which relationships develop inside close knit groups of friends. How do these friends pair off, form physical relations and get married or not.⁶ A further reason for interest in a fairly small network is that there is evidence that the maximum size of a meaningful network for interaction is around 50 or maybe up to 100 individuals.⁷ This means that when we analyze bigger networks (like the schools in the AddHealth) we are dealing with a very simplified version of the network in terms of what information we know about the interconnections or we are really studying a much larger group which is really an amorphous set of overlapping networks.

Various commentators have different lists of who was considered in the Bloomsbury Group and who was not. We will take a fairly eclectic view and include all those eleven individuals who most commentators agree were in the ‘core group’, namely: Clive Bell (CB), Vanessa Bell (VB), E.M. Forster (EMF), Roger Fry (RF), David ‘Bunny’ Garnett (BD), Duncan Grant (DG), John Maynard Keynes (JMK), Desmond McCarthy (DMC), Lytton Strachey (LS), Leonard Woolf (LW), Virginia Woolf (VW). Others, who are variously considered to be in the ‘outer group’ are: Thoby Stephen (TS), Saxon Sydney-Turner (SST), Adrian Stephen (AS), Gerald Brenan (GB), Dora Carrington (DC), Angelica Garnett (AG), Ottoline Morrell (OM), Ralph Partridge (RP), Harold Nicolson (HN), Vita Sackville-West (VSW), Mark Gerter (MG), Katherine Mansfield (KM), Lydia Lopokova (LL) and G E Moore (GEM). We present the simplest network figures of this group in two different years, at the beginning of their group in 1905 (Figure 1) and at the height of the group in 1925, Figure 2 – (using these personal named initials at the nodes). To establish a connection we deem two

⁶ See Verbrugge (1977), Dolton (1982) and McPherson et al (2001) for aspects of these questions.

⁷ See McCarthy et al (2001) or Hill and Dunbar (2003).

members of the group to be connected if they are: family, close friends, lovers, belong to the same group or College at university, become a member of the Omega artist co-operative or relate through the Hogarth Press.⁸ Using this figure we can identify the key player. Specifically it can be seen that the ‘key player’ in both 1905 and 1925, with the highest degree measure and the highest Katz-Bonacich centrality score is Lytton Strachey. In 1905 he is followed by Thoby Stephen, but by 1925 this has changed to Duncan Grant, John Maynard Keynes and Virginia Woolf. A detailed table of degree and highest Katz-Bonacich centrality score in both 1905 and 1925 is provided in Appendix A.

Using our data we can now explore whether there are any endogenous or exogenous social network effects of the network and whether there are any ‘key player effects’ i.e. by exploiting the Calvallo- Armengol et al (2009) result to examine the role of including the BC as a regressor in the outcome equation as a proxy for the extent to which being at the centre of the network shifts the outcome at the individual level.

There are many potential outcomes one could be interested in with this network but since this data is exploratory at this stage we seek only to establish the possible presence of network effects. A key event for this Bloomsbury Group was the onset of the First World War. This caused some (but not all) of the group to leave London and settle in Sussex. This was an exogenous event and each individual’s choice of what to do was likely to be influenced by the others in the group – hence giving rise to the possibility of network effects. This is what we model econometrically. So we use - the moving to Sussex (or spending a sizeable fraction of their time in Sussex – as the dependent variable. We use as regressors: gender, whether the person was an artist or literary and whether the person was bisexual or homosexual. The results – for various specifications are reported in Table 1. What we see is that being an artist or writer

⁸ We use all the available resources cited in the reference list relating to the Bloomsbury Group to determine these connections. The Adjacency matrix used for these calculations are available on request from the author.

is positively significant in this choice but no other x variable is. The network effects also show a consistent pattern – namely that the G_y and G_x term is never statistically significant but the Katz-Bonacich Centrality term is always significant at the 5% level. This suggests that being more central to the network correlates positively with the likelihood of moving to Sussex. These results are revealing in that they imply that the emphasis of the literature on the endogenous and exogenous network effects may be overplayed or not identified whereas the effects of a simply measure of the degree of centrality of the individual could be significant.

Table 1. Regression Results of Residence in Sussex.

	(1)	(2)	(3)	(4)	(5)
Artist or Literary	0.598** (3.42)	0.656* (3.05)	0.494** (2.98)	0.553** (2.69)	0.525** (2.53)
Male	-0.09 (-0.47)	-0.103 (-0.48)	-0.176 (-0.99)	-0.157 (-0.78)	-0.211 (-1.00)
Bisexual or Homosexual	0.072 (-0.42)	0.025 (-0.13)	-0.005 (-0.03)	-0.029 (-0.17)	-0.027 (-0.16)
Bonacich Centrality			0.754** (2.28)	0.732** (-2.06)	0.695** (-1.96)
G^*x		1.219 (-0.84)		0.594 (-0.43)	0.407 (-0.29)
G^*y		-1.262 (-0.86)		-0.704 (-0.51)	-0.825 (-0.59)
G^*BC					0.497 (-0.98)
Constant	0.132 (-0.58)	0.123 (-0.53)	-0.062 (-0.28)	-0.066 (-0.28)	-0.049 (-0.21)
R2	0.4	0.42	0.53	0.53	0.56
N	25	25	25	25	25

Detailed examination of the history of relationships in the Bloomsbury Group illustrate how complex networks may become. We have taken the simplest approach by graphing the network

at only two points in time. Nonetheless, this basic examination raises many questions. These questions help us to understand what we need to know to really understand networks and open up the possibility of seeing how the simple regression results reported above may be misleading – or indeed that some of this complexity may enable us to estimate more complete identified econometric models.

How are Networks Born and what are the Dynamics of Network Evolution: Most networks are not born fully formed. Mostly of them develop and change over time in subtle or gradual ways and some of them suddenly burst into life like a virus. In the case of the Bloomsbury Group the evolution was a gradual process – first based around a key family group of siblings – the Stephens –Virginia, Vanessa, Thoby and Adrian. Then when Thoby goes up to Trinity College, Cambridge University he meets up with fellow students Lytton Strachey, Leonard Woolf, Clive Bell and Saxon Sydney-Turner. Thoby joins the ‘Midnight’ society (at Trinity College) which has Lytton Strachey, Leonard Woolf and Clive Bell as other members. Lytton Strachey is also a member of the secret Cambridge society called the ‘Apostles’ and links with Keynes, E M Forster and G E Moore who are both at Kings College (along with Desmond McCarthy). Then at various points the family member links come into play: Grant is a cousin of Strachey and the Stephen’s: Vanessa, Virginia, Thoby and Adrian are siblings.

Homophily: To What Extent are Networks determined by the similarity of its members? Why was it that the Bloomsbury Group members gravitated together? What did they have in common? The answer is – a lot. They were all fairly wealthy and part of the educated aristocratic elite living primarily in London at the time. Most of them were educated at Cambridge University. Vanessa Bell, Roger Fry and Duncan Grant were artists who cemented their shared passion by setting up an artists cooperative called Omega Workshops Limited which operated from 1913-1919. Likewise, Strachey, Ralph Partridge, Leonard Woolf,

Virginia Woolf and Vita Sackville West all had literary futures – most of them cooperating in the founding in 1917 of the Hogarth Press under the direction of Leonard Woolf. In addition they nearly all had a shared view of individual freedom and morality largely inspired by the philosopher G E Moore.

Propinquity: Place and Spatial Dimensions: One network factor we seldom have information on is spatially how close members of a network are. In the case of the Bloomsbury Group we have a good insight into this. They lived for the most part – as the name of group suggests – in the heart of London’s fashionable Bloomsbury. But, as a direct result of the First World War, in 1916 key members – Vanessa Bell and Duncan Grant move to the Charleston farmhouse in Sussex and then most of the group came down for weekends, holidays and extended stays to Charleston. Only three years later in 1919 the Woolf’s moved very nearby to Monks House in Rodmell, Sussex. Keynes rented Tilton farmhouse which is next to Charleston in 1927 after many years of having one room in Charleston. Many weekends were spent walking across the South Downs between these locations. This means we have a good idea of how spatial location played a role in the unraveling of these relationships.

Relationships over different domains: There are many domains to the links between the members of the BG. There were membership links: because of family, links inside societies, literary links, artistic links, family connections, sexual liaisons, and a meeting of minds with respect to their philosophy, mores and outlook on society. Grant had sexual relations with: Garnett, Keynes, Strachey, Adrian Stephens and Vanessa Bell. Vanessa Bell had sexual relations with: Roger Fry, Clive Bell and Duncan Grant. Keynes had sexual relations with: Grant, Garnett, Strachey⁹ and Lydia Lopokova. Keynes and Strachey were both members of

⁹ There is some uncertainty about this relationship. Holroyd (1973) suggests that there was ‘a good deal more talk than action’ (see p. 244) but Davenport-Hines (2015) is clear in his assertion (see p.212).

the Apostle group of elite students at Cambridge; Grant was a lover of Keynes, Garnett, Bell, Strachey and many more; Bell, Fry and Grant were fellow artists; Forster, Virginia Woolf, Lytton Strachey and Vita Sackville-West were literary figures.

Non reciprocated relationships and links of different intensities:¹⁰ It was clear the Garnett had an unrequited love for Vanessa Bell and many members of the group had a much greater affection and strength of feeling towards Grant than Grant had for them. Ottoline Morrell had feelings for Virginia Woolf which were not reciprocated. Likewise, Dora Carrington was thwarted in her feelings for Lytton Strachey, Lytton Strachey for David Gerter, and David Gerter for Dora Carrington. It is reported that Lytton Strachey did not see eye to eye with Clive Bell¹¹. Vanessa Bell was married to Clive Bell but lived with Duncan Grant. Leonard Woolf was married to Virginia Woolf, and Harold Nicolson was married to Vita Sackville-West but it was Vita and Virginia who fell in love with each other. In other elements of their interrelationships there were other asymmetries. For example, Keynes subsidized the running of Charleston for years, and financially supported other members of the group – for example by making Grant an annuity from 1937 onwards and paying school fees for Garnett's children.

12

4. The Size and Scope of Networks: Some Evidence from the AddHealth Data.

On the face of it, one of the simplest but most fundamental questions about networks that one can pose is – how widespread is a network in terms of its connections? Literally how many people have connections with how many others in the network? In many situations we would wish to know what is the distribution of nominated friends or contacts in any given population.

¹⁰ Technically, the G matrix can be rescaled to allow for intensity of links in the network. Subject to row normalization the results in the previous section go through. See DeGroot (1974).

¹¹ See Holroyd (1973) p

¹² See Spalding (1998), p.351.

In reality how many contacts does each person have – and what is the distribution of the number of connections in a given population? In more detail we can ask the question of whether there is a difference between the number of connections that an individual says they have as ‘friends’ (out-degree), and the number of friends who actually declare a person as a friend (in-degree). The sum of in-degree and out-degree is the (total) degree. The extent of network connections which are reciprocated and ‘non-reciprocated’ could be quite important in certain contexts.

These basic descriptive statistics are important for many health, epidemiology and economic processes which involve diffusion and contagion processes. This may in turn have implications for modeling public policy interventions. Consider a simple example that routinely concerns epidemiologists. Suppose there is an outbreak of a virulent communicable disease spread by close contact between people and that we need to get an idea of how quickly this might spread and how many people might be affected. Here, knowledge of friendship networks may be extremely helpful in making those predictions and developing public health intervention strategies. This all assumes we can model how widespread the friendship is.

The amazing reality is that there are simply not many datasets in the world that can answer these basic questions. One of the best datasets we have is the AddHealth (the National Longitudinal Study of Adolescent to Adult Health) in the USA which is collected out of the University of North Carolina. The survey data started in 1994 (WAVE I) with a survey of a large number of 15 year olds across the USA. Wave II was collected a year later when they were 16 with WAVE III following when they were 22 and WAVE IV following in 2008 when they were 28. The survey collected large amounts of information about the health, social and economic circumstances of the family and the educational and early labour market outcomes of the respondents. The most remarkable aspect of the data is that High School friendship information was collected in the first two waves. This took the form of the respondents being

asked to nominate their 5 best male and 5 best female friends. It is this data which makes this dataset nearly virtually unique for the study of networks and their implications.

Most simple characterizations of the network formation process (see Jackson 2009, p11-13) assume the link between any two nodes forms independently with a specific probability. Such a randomly generated network gives rise to a Poisson degree distribution. Another alternative common assumption in the literature is the ‘Scale Free’ distributional assumption. This degree distribution has no underlying stochastic model of friendship formation. The problem with both these distributions is that they have very limited empirical applicability to the real world data. Dolton et al (2016) show empirically is that it is unlikely that either the Poisson or Scale Free distribution is likely to be observed in reality with empirical data. They examine the distribution of friends in the AddHealth data and show that they are not Poisson or Scale Free distributed. We reproduce some of the empirical descriptive statistics in this paper to aid our basic understanding of what a friendship network looks like. Figure 2 below shows this in-degree distribution of friends in the AddHealth data. Figure 3 graphs the out-degree distribution, Figure 4 the total degree distribution, Figure 5 the distribution of reciprocated friendship nominations and, finally, Figure 6 graphs the distribution of the Katz-Bonacich Centrality measure across the people in the network. The paper then goes on to show how it is possible to generate a stochastic model of friendship formation by assuming heterogeneity which takes a Gamma form. We also show that the form of this heterogeneity could be endogenous and related to a form of contagion process. It turns out that the distribution which provides the closest fit the data is a Negative Binomial distribution. This result could have important implications for many epidemiological and other applications. Specifically, if we take the examples below we see that the actual empirical model gives rise to an underlying distribution on both in-degree (Figure 2), out-degree (Figure 3) and degree (Figure 4) are all negatively

skewed and the Negative Binomial distribution fits the data much better than the best Poisson model which could approximate the data. Looking at the degree distribution in Figure 4 the modal number of nominations is 4. In contrast, in the Poisson model with the same mean we see the model number of friends is 6. This is a huge difference and could have major implications for modeling the consequences of friendship distributions. Let's take our example – if each person has had close affiliation with 6 people rather than 4 people the consequences for the epidemiological spread of a communicable disease is rather momentous.

Figure 2. The In-Degree Distribution In AddHealth Data.

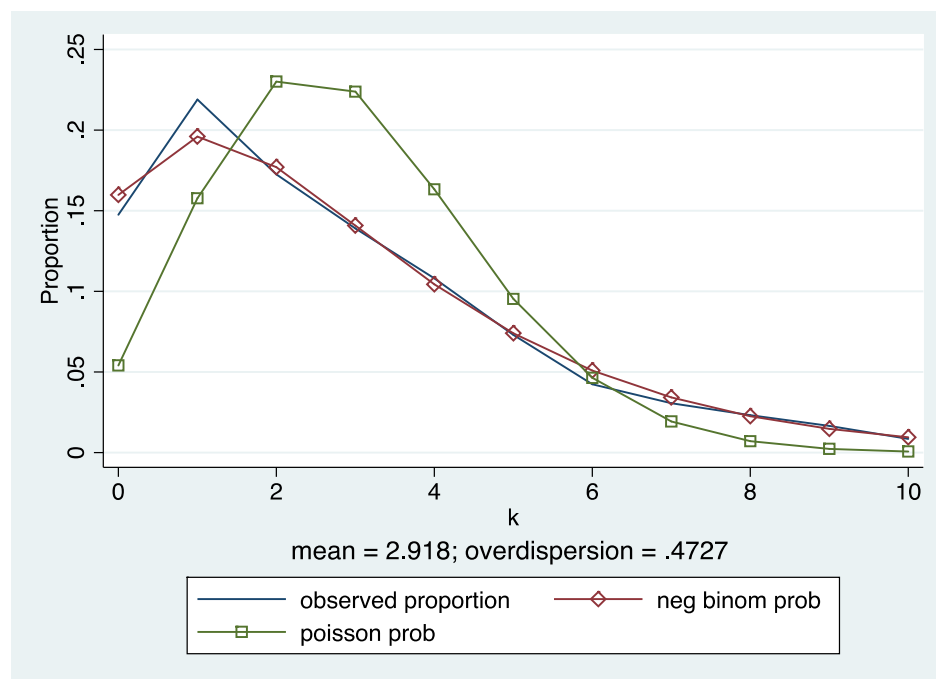


Figure 3. The Out-Degree Distribution In AddHealth Data.

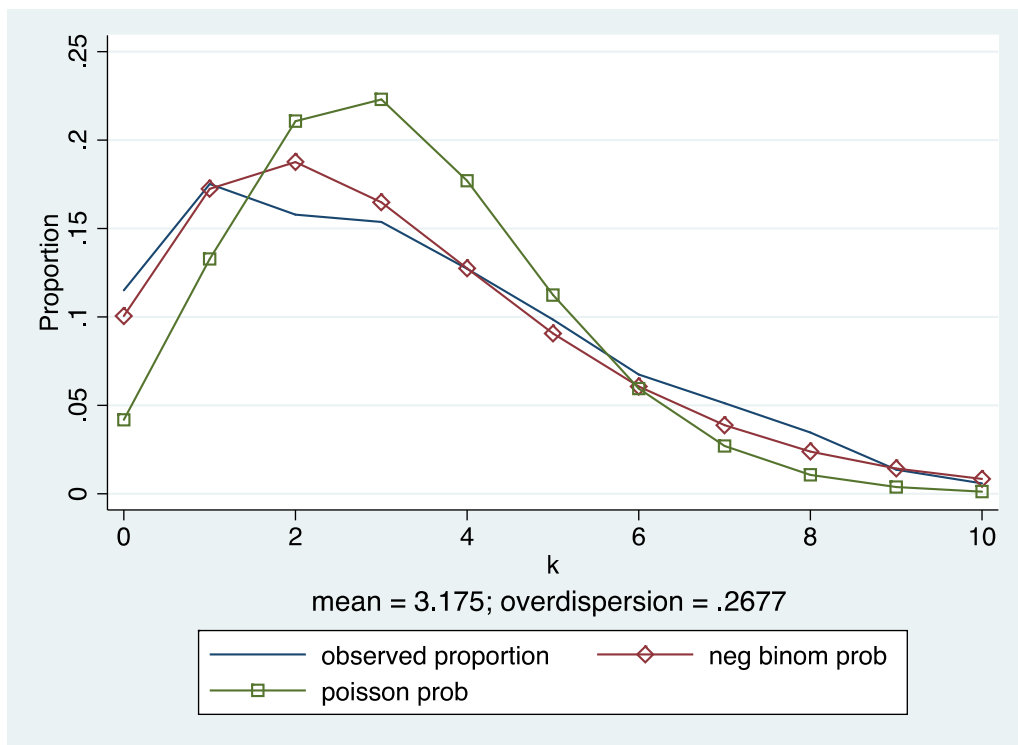
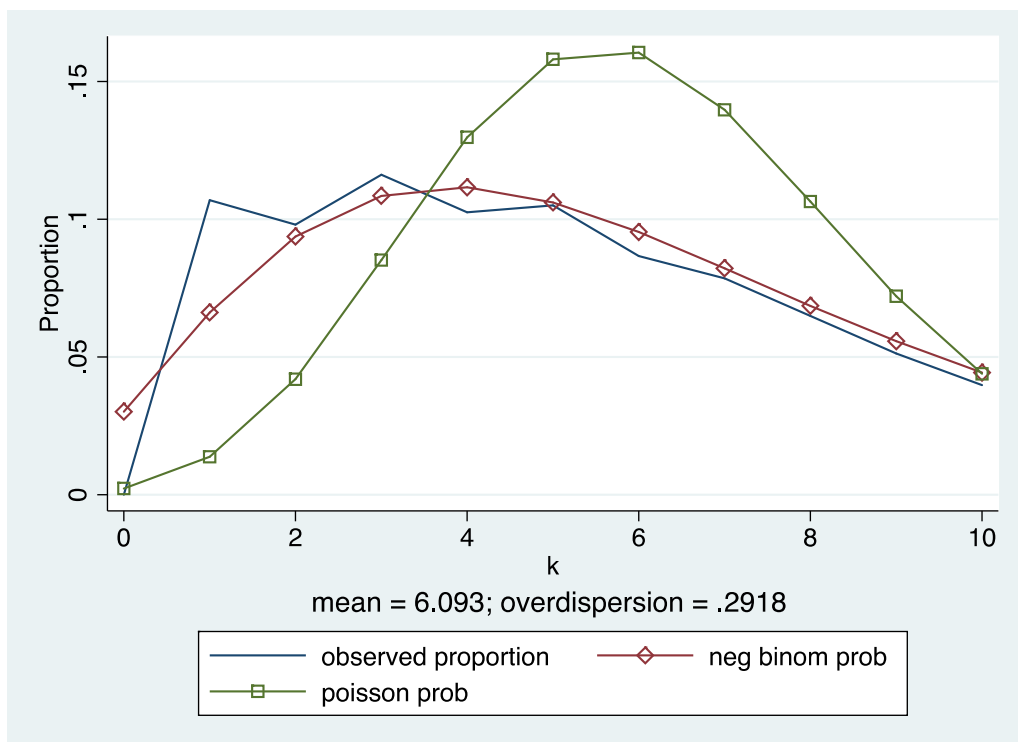


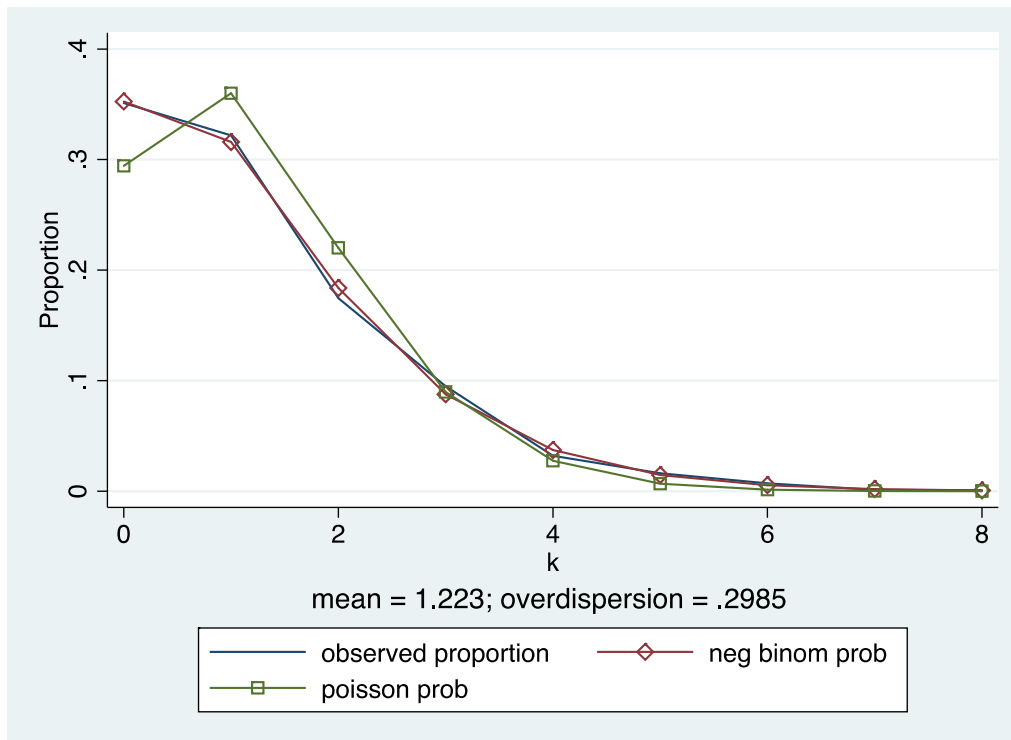
Figure 4. The Degree Distribution In AddHealth Data.



It is only when we move to consider the distribution of reciprocated friendship nominations (which has a modal value of 1) that we see a fairly close concordance between both the Negative

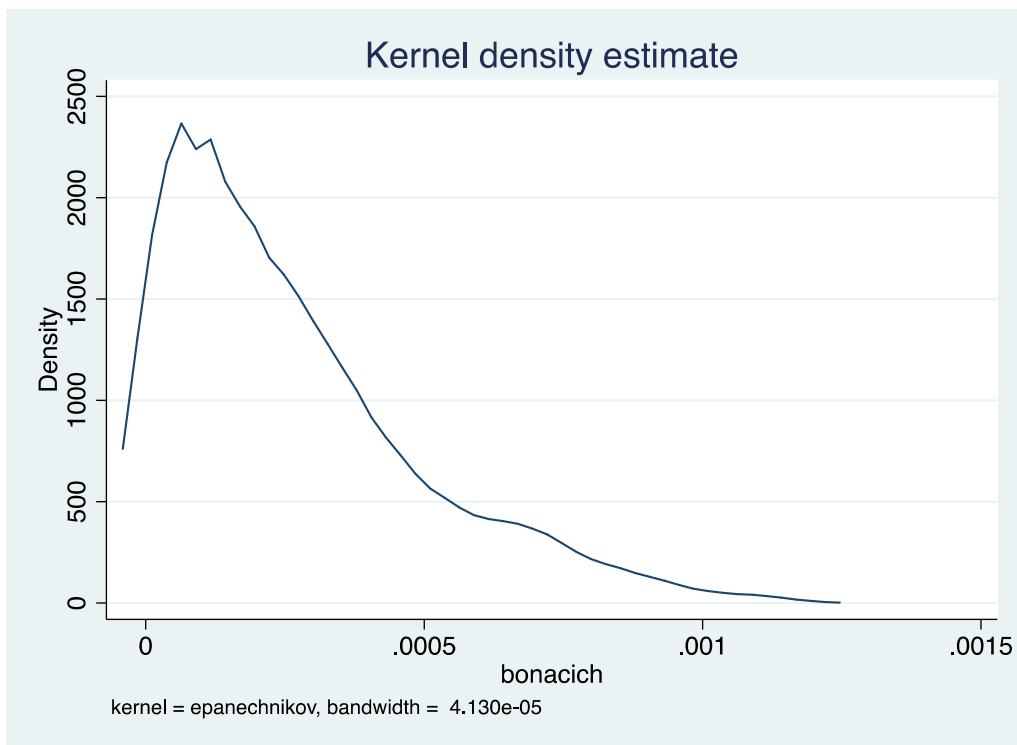
Binomial and the Poisson in empirically fitting the data – although, even here the Negative Binomial model does better.

Figure 5. The Number of Reciprocated Nominations Distribution In AddHealth Data



In many key contributions to the literature the role of centrality measures in a network assume a central importance. In Figure 6 we graph the kernel distribution plot of the Katz-Bonacich centrality measure. We see that it is negatively skewed with a long right tail – at present economic theory remains silent over what the distribution of this measure should be.

Figure 6. The Katz-Bonacich Centrality Distribution In AddHealth Data.



Many other challenges remain for the economic theory of networks to be compatible with the empirical nature of what we observe with network data. Most of our econometric models are based on the assumption that all friendship nominations are reciprocated e.g. Calvo-Armengol et al (2009) but this is clearly not true. In addition, other empirical work is based on limited numbers of friendship nominations with the assumption that there is no correlation between in-degree and out-degree, eg see Conti et al (2013). The importance of these difference between the theoretical models and the econometric estimation are worthy of attention – as are their implications for practical policy implementations.

5. The Challenges for Empirical Network Studies.

There are many challenges for empirical work relating to social and economic networks. This

paper has described some of them by examining two datasets. The AddHealth was used to question our underlying understanding of how networks are formed. We showed how the standard model of random meeting, with a set probability of forming friendships does not yield a distribution on degree that matches the empirical data. We suggested that the distribution of in-degree, out-degree, degree and degree with reciprocity could all be subtly different. In addition, we drew attention to the fact that many of the theoretical contributions assume that the networks under consideration all have reciprocated friendship nominations. We showed that this was not true. Under these circumstances it is not clear what the consequences of the violation of this assumption might be.

The paper also describes how the Bloomsbury Group was formed and transformed over time. Its detailed analysis showed how the network changes in its structure. Many people will be surprised to learn that Lytton Strachey turns out to be the most important key player in both 1905 and 1925. Although some of the other important figures will be less of a surprise by 1925 – including Duncan Grant and Vanessa Bell. We used our data to examine the determinants of spending a substantial amount of time in Sussex after the onset of World War I in 1914. We find that how central you are to the network plays a significant role in the determination of this dependent variable. In contrast we found that the endogenous and exogenous network effects were not significant. One interpretation which is in line with the theoretical identification results in section 2 is that these effects are actually hard to identify. Our conclusions concur with the findings we get when we use AddHealth data to model the determination of earnings for High School graduates 13 years after leaving school in the labour market (See Barbone and Dolton, 2016). Namely, that it is the impact of key player effects which is robust and that exogenous and endogenous network effects are not consistently significant. Despite these interesting findings our simple approach is only illustrative and must lead us to address other

important questions regarding how networks function and what the nature of their effects are.

An important area of concern is the endogeneity of the adjacency matrix. How is it possible to identify what determines the formation of friendships separately from what determines outcomes? Are there any circumstances in which homophily or propinquity are exogenous and can be used as IVs in the estimation of an endogenous G matrix?

At the more detailed level we can learn from the Bloomsbury example that we need to be concerned about: the intensity and dimensions of relationship links, can a link between two individuals that works on different levels be modeled as the sum of these different effects? We also need to potentially address the problems of un-reciprocated asymmetric sentiments between friends. Can friendships that turn to animosities be treated with negative numbers in an adjacency matrix?

When we move from a network where we can know a lot about the characters in the network – like the Bloomsbury Group to a situation like the AddHealth data where we know only a limited amount about a lot of people – how do we reconcile the econometric methods. In the AddHealth the pupils were asked to name only 5 boy friends and 5 girl friends. How much are we missing if the friendship network of a typical girl is 15 girl friends and no male friends? Another issue is how we treat partially sampled networks. What is missed for the pupils who have no friends inside school but lots of friends outside. We have focused our simple distributional plots on only those schools which were ‘saturated’ – i.e. all the pupils were sampled and asked about their friends. But even here the distribution of out-degree will be somewhat truncated as by definition we only see these nominations if they are for other pupils in the school. To the extent that the nominations include friends outside school then we truncate the number of possible friends in the out-degree distribution. The situation would have become much worse if we had analyzed the partially sampled schools where only a random sample of

pupils where asked to provide data on friendship nominations. We concur with Chandrasekhar (2016) when he suggests that we cannot infer much about the nature of the network or its impacts from such data. The position could be worse still if the data used has a partial sample and also a very limited number of friendship nominations – as in the case of the Conti et al (2013) paper – where respondents were asked to nominate only up to three friends. In general, in studying a network - how far do we need to take a search along the paths of the network to be able to model the endogenous or exogenous effects of the network on an individual's outcome.

If we do have to treat G as exogenous - under what circumstances is the Calvo-Armengol et al (2009) assumption of a quadratic effort in network formation adequate? What might be the biases of ignoring network connections if we estimate a model with no network effects when the data of the model have clear relationships between the rows of the data matrix? Specifically, what are the consequences of mistakenly assuming cross section independence on our estimation and inference? More complex still, are the estimation of the biases of running a Linear-in-Means type, peer effects model and assuming that the friendship nominations of peer groups are exogenous? Which of these two biases are worse? We do not know.

A positive direction that the empirical research on networks could take is to consider modeling the friendship formation process mechanism endogenously. Papers like Goldsmith-Pinkham and Imbens (2013) do this explicitly identifying the model by joint parametric unobserved heterogeneity assumptions. This hugely computationally expensive approach resulted in only limited success with little evidence of the presence of network effects – even only in a very simple model of educational attainment in the AddHealth data.

What are the consequences of these observations for the future on improved econometric estimation of network effects? There are six major implications. Firstly, we need to consider

collecting much more complete network information of a saturated kind. Secondly we need to consider collecting this data repeatedly at regular intervals to understand how friendships form and change and how networks are modified as a result. Thirdly, we should find ways of collecting network data on the dimensions that agents interact – are they fellow students in the same clubs, lovers, friends, fellow literati, fellow artists, or what? Fourthly, in the effort to identify our econometrics models, we need to consider carefully the collection of other data which may help us to understand the mechanisms of how networks form – I am thinking specifically of collecting exogenous good data on what factors may influence how networks form – in the case of the work by Barbone and Dolton (2016) we attempted to use the spatial proximity of the pupils homes and the distance between them to establish the extent to which this might mean they meet up more frequently on the bus to school – the ‘Yellow Bus’ effect. Such information may provide useful IV variables that relate to the formation of friendships but are independent of the unobserved heterogeneity in outcomes in later life. Fifthly, we need to consider how experimental economics may take up the challenge of helping to explain network formation and its consequences – for example by running an experiment with students, who at the beginning of the academic year do not know each other, but through random class assignment get to know a certain group of people. We could then subject them to routine laboratory experiments to measure the importance of networks. Finally, it is clear from our descriptive statistics relating to the distortions of the key network variables that there is a mismatch between the basic theories of how networks are formed and their predictions relating to the likely empirical distributions we would expect to see in the data.

Perhaps the most illuminating suggestion in this paper is that basic measure of Katz-Bonacich Centrality may effectively capture the basic nature of network effects. We found that a significant ‘key player’ effect even in a small sample network like the Bloomsbury Group. This

finding concurs with the importance of the key player effect in the AddHealth data relating to earnings 13 years after High School graduation. This suggests that maybe the emphasis of the empirical network literature on trying to identify endogenous and exogenous peer type effects of networks could be slightly misplaced. Instead a simple characterization that controls for how central a person is in a network could be more instructive and insightful provided we have good quality network data.

References.

- Advani, A. and Malde, B. (2014) 'Empirical Methods for Networks Data: Social Effects, Network Formation and Measurement Error', IFS Working Paper W14/34.
- Angrist, J. D. (2014), 'The Perils of Peer Effects', *Labour Economics* 30, 98-108.
- Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.
- Barabasi, A-L. (2002) *Linked*, Perseus Publishing.
- Barbone, L and Dolton, P.J. (2016) 'Key Player Effects: The High School Networking Impact on Subsequent', University of Sussex, mimeo.
- Bell, Q. (1968) *Bloomsbury*, Giant Phoenix, London.
- Bell, Q. (1995) *Bloomsbury Revealed*, Columbia University Press, New York.
- Borgatti, S, Everitt, M. and Johnson, J. (2013) *Analyzing Social Networks*, SAGE.
- Bloch, F. and Dutta, B. (2011) 'Formation of Networks and Coalitions', chp 16 in *Handbook of Social Economics*, vol 1A. North Holland.
- Blume, L, Brock, W, Durlauf, S. and Ioannides, Y. (2011) 'Identification of Social Interactions', chp 18 in *Handbook of Social Economics*, vol 1B. North Holland.
- Bonacich, P. (1987), 'Power and Centrality: A Family of Measures', *American Journal of Sociology* 92(5), 1170-1182.
- Boucher, V. and Fortin, B. (2016) 'Some Challenges in the Empirics of the Effects of Networks', in Bramoulle Y, Galeotti, A. and Rogers, B. *Oxford Handbook of the Economics of Networks*, Oxford University Press.
- Bramoulle, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* 150(1), 41-55.
- Calvo-Armengol, A.Patacchini, E. and Zenou, Y. (2009), 'Peer Effects and Social Networks in Education', *Review of Economic Studies* 76(4), 1239-1267.
- Chandrasekhar, A. (2016) 'Econometrics of Network Formations' in Bramoulle Y, Galeotti, A. and Rogers, B. *Oxford Handbook of the Economics of Networks*, Oxford University Press.
- Christakis, N. and Fowler, J. (2007) 'The Spread of Obesity in a large Social Network over 32 Years', *New England Journal of Medicine*, 357, 370-9.
- Christakis, N. and Fowler, J. (2008) 'The Collective Dynamics of Smoking in a Large Social Network', *New England Journal of Medicine*, 358, 2249-2258.
- Christakis, N. and Fowler, J. (2011) *Connected*, Harper Press.
- Coleman, J.S., E. Katz, and H. Menzel (1966) *Medical Innovation: A Diffusion Study*, Indianapolis, Ind.: Bobbs-Merrill.
- Conti, G., Galeotti, A., Muller, G. and Pudney, S. (2013), 'Popularity', *Journal of Human Resources* 48(4), 1072-94.
- Currarini S, Jackson M.O., Pin P. (2009) "An economic model of friendship: homophily, minorities and segregation," *Econometrica*, Vol. 77, No. 4, 1003-1045,
- Davenport-Hines, R. (2015) *Universal Man: The Seven Lives of John Maynard Keynes*, William Collins, London.
- De Giorgi, G, Pellizzari, M and Redaelli, S. (2010) 'Identification of Social Interactions through Partially Overlapping Peer Groups', *American Economic Journal: Applied Economics*, 2, 241-275.
- De Groot, M. H. (1974) 'Reaching a Concensus', *Journal of the American Statistical Association*, 69:118-121.
- De Weerd J. 2004. *Risk-sharing and Endogenous Network Formation*. In *Insurance Against Poverty*, ed. S Dercon. Oxford: Oxford Univ. Press
- Dolton, P. (1982) 'The 'Marriage Game': An Assignment Problem with Indivisibilities', *Theory and Decision*, 14, 373-389.
- Dolton, P. M. Stops and C. Rosazza-Bondibene (2016) 'Identifying the Employment Effect of Invoking and Changing the Minimum Wage' *Labour Economics*. Vol.37, 54-76, 2015.
- Dolton, P, Kiraly, F. and Barbone, L. (2016) 'Reconciling the Distribution of Network Connections', University of Sussex, mimeo.

- Easley, D. and Kleinberg, J. (2010) *Networks, Crowds and Markets*. Cambridge University Press.
- Edel, L. (1979) *Bloomsbury: A House of Lions*, Penguin Books Ltd, Harmondsworth.
- Elliott, M.L., B. Golub and M.O. Jackson (2012) "Financial Networks and Contagion," SSRN working paper 2175056
- Fafchamps M. and Lund S. 2003. Risk-sharing Networks in Rural Philippines. *J. Dev. Econ.* 71: 261 – 87
- Fryer, R. (2007) "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century," *Journal of Economic Perspectives* 21(2):71–90.
- Glaeser E, Sacerdote B, Scheinkman J. 1996. Crime and social interactions. *Quarterly Journal Economics*. 111: 507 – 48
- Goyal, S. (2007) *Connections: An Introduction to the Economics of Networks*, Princeton University Press.
- Goyal, S. (2011) Learning in Networks, chp 15 in *Handbook of Social Economics*, vol 1A. North Holland.
- Goldsmith-Pinkham, P. & Imbens, G. W. (2013), 'Social Networks and the Identification of Peer Effects', *Journal of Business & Economic Statistics* 31(3), 253-264.
- Graham, B.S. (2008) 'Identifying social interactions through conditional variance restrictions', *Econometrica*, 96, 485-494.
- Granovetter M. 1973. The strength of weak ties. *Am. J. Sociol.* 78: 1360 – 80
- Griliches, Z. (1957) "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica* 25(4):501–522.
- Hill, R and Dunbar, I. 'Social Network Size in Humans', *Human Nature*, 14:1, 65.
- Holroyd, M. (1973) *Lytton Strachey*, Heineman, London.
- Hsieh, C-S and Lee, L.F., (2016) 'A Social Interactions Model with Endogenous Friendship Formation and Selectivity', *Journal of Applied Econometrics*, 31, 301-319.
- Ioannides YM, Datcher-Loury L. 2004. Job information networks, neighborhood effects and inequality. *J. Econ. Lit.* 424: 1056 – 93
- Jackson, M.O. (2008) *Social and Economic Networks*, Princeton University Press
- Jackson, M. O. (2014), 'Networks in the Understanding of Economic Behaviors', *Journal of Economic Perspectives* 28(4), 3{22.
- Jackson, M. (2011) An Overview of Social Networks and Economic Applications, chp 12 in *Handbook of Social Economics*, vol 1A North Holland.
- Kadushin, C. (2012) *Understanding Social Networks: Theories, Concepts and Findings*. Oxford University Press.
- Knoke, D and Yong, S. (2008) *Social Network Analysis*, 2nd edition. SAGE Publications.
- Manski, C. (1993) 'Identification of Endogenous Social Effects: The Reflection Problem', *Review of Economic Studies*, 60(3), 531-542.
- McCarthy et al (2001) 'Comparing Two Methods of Estimating Network Size', *Human Organization* 60:1, 32.
- McPherson, M., L. Smith-Lovin, and J.M. Cook (2001) "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* 27:415–444
- Myers C.A., and G.P. Shultz (1951) *The Dynamics of a Labor Market*, New York: Prentice-Hall.
- Newman, M. (2010) *Networks: An Introduction*, Oxford University Press.
- O'Malley, A. and Marsden, P. (2008) The Analysis of Social Networks, *Health Service Outcomes Research Methods*, 8, 222-269.
- Padgett, J.F., and C.K. Ansell (1993) "Robust Action and the Rise of the Medici, 1400–1434," *American Journal of Sociology* 98:1259–1319.
- Pentland, A. (2014) *Social Physics: How Social Networks Can Make You Smarter*, Penguin.
- Rainie, L. and Wellman, B. (2012) *Networked: The New Social Operating System*. MIT Press.
- Rees, A. J., Shultz GP. 1970. *Workers in an Urban Labor Market*. Chicago: Univ. Chicago Press
- Reiss AJ. 1988. Co-offending and criminal careers. In *Crime and Justice: A Review of Research*, Vol. 10, ed. M Tonry. Chicago: Univ. Chicago Press
- Spalding, F. (1998) *Duncan Grant: A Biography*, Pimlico, London.
- Spalding, F. (2005) *The Bloomsbury Group*, National Portrait Gallery, London.

Uzzi B. 1996. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *Am. Sociol. Rev.* 61: 674-98

Vega-Redondo, F. (2007) *Complex Social Networks*, Cambridge University Press.

Verbrugge, L.M (1977) The Structure of Adult Friendship Choices, *Social Forces*, 56:2, 576-597

Weisbuch, G., A. Kirman, and D. Herreiner (2000) "Market Organization," *Economica* 110:411–436.

Wasserman, S and Faust, K. (1994) *Social Network Analysis*, Cambridge University Press.

Table A1 - Summary Statistics

Name	deg1905	bc1905	deg1925	bc1925
Clive Bell	3	0.295	4	0.4884268
Vanessa Bell	3	0.383	8	0.78718563
E M Forster	4	0.246	4	0.2448263
Roger Fry	3	0.145	5	0.43504632
David 'Bunny' Garnett	0	0.000	3	0.4827811
Duncan Grant	1	0.295	8	0.75196728
J M Keynes	5	0.246	6	0.53724738
Desmond McCarthy	0	0.188	0	0.22811369
Lytton Strachey	6	1.000	11	1
Leonard Woolf	1	0.776	5	0.81885706
Virginia Woolf	3	0.113	7	0.73827698
Thoby Stephen	5	0.799	5	0.74961175
Saxon Sydney-Turner	3	0.532	3	0.32791376
Adrian Stephen	2	0.383	4	0.65781404
Gerald Brenan	0	0.000	1	0.07919658
Dora Carrington	0	0.000	6	0.42256009
Angelica Garnett	0	0.000	3	0.37895264
Ottoline Morrell	0	0.000	2	0.26824475
Ralph Partridge	0	0.000	5	0.6383856
Harold Nicolson	0	0.000	1	0.07992873
Vita Sackville-West	0	0.000	4	0.42646652
Mark Gerter	0	0.000	2	0.26661745
Katherine Mansfield	0	0.000	3	0.27040561
Lydia Lopokova	0	0.000	1	0.10069137
G E Moore	3	0.441	3	0.33399781